Statistical Learning with Time-varying Parameters*

by Bruce McGough

Oregon State University, October 17, 2000

Abstract

In their landmark paper [1], Bray and Savin note that the constant parameters model used by their agents to form expectations is misspecified and that, using standard econometric techniques, agents may be able to determine the time-varying nature of the model's parameters. Here, we consider the same type of model as employed by Bray and Savin except that our agents form expectations using a perceived model with parameters which vary with time. We assume agents use the Kalman filter to form estimates of these time-varying parameters. We find that, under certain restrictions on the structure of the stochastic process and on the value of the stability parameter, the model will converge to its rational expectations equilibrium. Further, the restrictions on the stability parameter required for convergence are identical to those found by Bray and Savin.

1 Introduction

Modern stochastic macroeconomic models typically include, among the factors governing their dynamics, dependence upon the predicted values of endogenous variables. The standard method of analysis of such models comes from the theory of rational expectations. According to this theory, economic agents are assumed to form predictions using conditional mathematical expectations; when these conditional expectations are formed with respect to the distributions of the actual stochastic processes generating the data, the economy is said to be in a rational expectations equilibrium (REE). This notion of equilibrium is well established as the discipline's benchmark; however, it is not without criticism. In [3], Evans and Honkapohja note on page 453,

*I would like to thank George Evans for many valuable discussions. All errors are mine.

> rational expectations . . . assumes that agents know the true economic model generating the data and implicitly assumes coordination of expectations by the agents.

Further, examples of macroeconomic models with multiple rational expectations equilibria are abundant; the theory gives no indication as to which equilibrium is likely to govern the behavior of the economy. To address these criticisms, some economists choose to weaken the notion of rationality. Instead of assuming agents know the true economic model generating the data, researchers assume agents are boundedly rational. The manifestation of this assumption in models of statistical learning is that agents know the structure of the rational expectations equilibrium and form estimates of the relevant parameters adaptively, using statistical algorithms. With these estimates, agents form their expectations of the values of the endogenous variables. One can then consider whether the economy eventually approaches a rational expectations equilibrium, that is, do the parameter estimates eventually converge (in some probabilistic sense) to the corresponding rational expectations equilibrium parameter values.

The first authors to consider such a model were Bray and Savin. In their landmark paper [1], they showed that if the economy is governed by a simple cobweb model, and if agents have a perceived model of the same linear functional form as the rational expectations equilibrium, and if agents estimate the parameters of this model using ordinary least squares, then the estimates indeed converge to the associated REE values, for appropriate values of the stability parameter. Their method of proof was quite technical and based on the theory of Martingales. Marcet and Sargent, [7], were able to extend this result to more general linear models using Ljung's theory on recursive stochastic algorithms, [6]. Evans and Honkapohja derived a Ljung type result designed specifically for application to economic models and subsequently extended these convergence results to multivariate linear models. See [4] for details. Much further work has been done. For a survey and brief history see [3].

All the results mentioned in the previous paragraph are derived assuming agents use OLS as their method of estimation. Implicit, then, is the presumption by agents that the parameters of the model are constant. However, since the estimates themselves necessarily affect the true values of the parameters, these constant parameter beliefs by agents are erroneous. Bray and Savin knew this to be a concern and used simulations to consider whether agents could detect the time-varying nature of the parameters. They found that for certain initial conditions and parameter values, agents may in fact determine that the model is misspecified. Since it may be possible for agents to realize that the parameters are not constant, it is important to analyze models in which agents believe the parameter values vary with time, thus the topic of this paper.

The behavior of an economic model based on agents with time-varying parameter beliefs has been considered by Jim Bullard, [2]. Bullard uses a general linear reduced form model, of which the above mentioned cobweb model is a special case, to show that if agents believe the parameters of the perceived model follow a random walk with *i.i.d.* noise term, then the economy never converges to the REE. This result is not surprising: if agents believe the parameters of the economy will not settle down, then their estimates of these parameters will not settle down because the agents will always attribute some of the noise in the model to movement in the parameter values. We conclude that for convergence to a rational expectations equilibrium to occur, the agents must believe the conditional variance of the time-varying parameters decreases to zero.[1] And this is a natural assumption for the agents to make. In particular, if agents initially use OLS to form their estimates, then the results of Bray and Savin tell us that the conditional variance of the process describing the actual parameters will decrease to zero.[2]

In this paper, we analyze the asymptotic behavior of an economy described by a simple cobweb model with agents who have time-varying parameter beliefs. We find that if they believe the parameters of the economy follow a random walk, and if the conditional variance of this random walk decreases rapidly enough, then convergence to the rational expectations equilibrium obtains for appropriate values of the stability parameter.

This paper is organized as follows: Section 2 begins with a review of the simple cobweb model and results of Bray and Savin and then presents the modification of the model which allows for time-varying parameter beliefs. A change of variables is presented which allows for simpler analysis of the stochastic processes. The main result of the paper ends the section. In Section 3 a more general cobweb model is considered and tools from the theory of stochastic approximation are used to show convergence in this case. A connection with E-stability is also discussed. Section 4 concludes. Most of the technical proofs are relegated to the Appendices.

[1] This conclusion is explained more fully in Section 2 below.

[2] Margaritis, [8] also considers a model with time-varying parameters. For convergence to a point to obtain, his results require that the gain of the adaptive algorithm tends to zero. As we note in Section 2.2.2, and as was shown by Bullard in [2], this can not hold in our model if the conditional variance of the time-varying parameters is positive definite.

2 Bray and Savin's Cobweb Model

2.1 Constant Parameter Beliefs

In this section, we consider the same cobweb model as analyzed by Bray and Savin. The reduced form of this model is[3]

$$y_t = x_t' m + a E_t^* y_t + \nu_t, \tag{1}$$

where y_t is the endogenous variable, $x_t \in \mathrm{R}^n$ is an exogenous *i.i.d.* process observed at time t, the first component of which is 1, ν_t is an unobserved white noise shock, and $E_t^* y_t$ is the agents' expectation of the value of y_t formed using information up to and including time t. The model is closed by specifying the form of the expectations operator. Provided agents behavior rationally,

$$E_t^* y_t = E\left(y_t | \Omega_t\right),$$

where Ω_t represents the agents' information set. The unique rational expectations equilibrium, that is, the final form of the model consistent with the assumption of rationality, is then easily computed to be

$$y_t = x_t' \left(\frac{m}{1-a}\right) + \nu_t.$$

To weaken the assumption of rationality and subsequently incorporate learning into their model, Bray and Savin postulate agents believe

$$y_t = x_t' \beta + \varepsilon_t$$

but are unaware of the value of β. Further, Bray and Savin assume their agents use OLS to estimate β and then use this estimate to form their expectations. Specifically, let b_t be the OLS estimate of β using data $(x_1, y_1), \cdots, (x_t, y_t)$. Then $E_t^* y_t = x_t' b_{t-1}$. Agents' expectations feed back into the reduced form model, (1), to yield the actual data generating process

$$y_t = x_t'(m + a b_{t-1}) + \nu_t. \tag{2}$$

Notice the parameter modifying x_t, namely $(m + a b_{t-1})$, is time dependant, contrary to the assumption of the agents.

To complete their analysis, Bray and Savin use recursive least squares, together with true process (2), to write the sequence of estimators, b_t, as

$$b_t = (I + (a-1)\frac{1}{t} V_t x_t x_t') b_{t-1} + \frac{1}{t} V_t x_t x_t' m + \frac{1}{t} V_t x_t u_t,$$

[3] For a careful derivation of this reduced form, see [1].

where $V_t = t\left(\sum_{i=1}^{t} x_i x_i'\right)^{-1}$. Their main result is as follows.

Theorem 1 (Bray and Savin, 1986) *If* $a < 1$ *then* $b_t \to \frac{m}{1-a}$ *almost surely.*

It is important for our work to observe that the proof of this theorem does not rely on the structure of the process V_t, but only that it converges to $(Ex_t x_t')^{-1}$ almost surely.

2.2 Time-Varying Parameter Beliefs

In this section we alter the model of Bray and Savin by allowing the parameters of the agents' perceived model to vary with time. We then attempt to analyze the resulting asymptotic behavior of the economy. This analysis requires imposing a structure on the believed process describing the time-varying parameters. Here we consider the process to be a random walk with potentially variable conditional variance. The reasons for choosing a random walk are fourfold: first, a random walk is a standard model of time-varying parameters; second, it is consistent with the learning literature, see, for example, [2]; third, if the conditional variance is zero then the random walk reduces to the constant parameter model considered by Bray and Savin; and fourth, its simplicity allows for analytic tractability.

We modify Bray and Savin's model as follows. Assume agents believe

$$\begin{aligned} y_t &= \beta_t' x_t + \epsilon_t \\ \beta_{t+1} &= \beta_t + \eta_t \end{aligned}$$

and that x_t and β_t are independent. We assume $var(\eta_t)$ is positive definite. Denote by b_t the estimate of β_{t+1} using data available at time t. Then $E_t y_t = b_{t-1}' x_t$.[4] Inserting this into equation (1) yields the true data generating process

$$y_t = (m + \alpha b_{t-1})' x_t + u_t. \tag{3}$$

Given agents' beliefs, the Kalman filter is a natural estimator for b_t[5]. The recursions for the filter

[4] Agents are free to use x_t in their estimation of b_{t-1} but the independence assumption made above implies that the realization of x_t will not alter their estimates.

[5] If the x_t and u_t are normal, it is the optimal estimator.

are given by

$$K_t = P_{t-1}[\sigma^2 + x_t' P_{t-1} x_t]^{-1} \tag{4}$$

$$b_t = b_{t-1} + K_t x_t [y_t - b_{t-1}' x_t] \tag{5}$$

$$P_t = P_{t-1} - K_t x_t x_t' P_{t-1} + var(\eta_t) \tag{6}$$

where y_t is given by (3). Like Bray and Savin, our goal is to analyze the process b_t and determine under what conditions it converges to $\frac{m}{1-a}$.

We first prove a few results concerning the process P_t. Because it will be used repeatedly, we state the well known matrix inversion lemma here.

Lemma 2 *Let W, X, Y, Z be conformable matrices. Then, provided the indicated inverses exist,*

$$[W + XYZ]^{-1} = W^{-1} - W^{-1}X[ZW^{-1}X + Y^{-1}]^{-1}ZW^{-1}.$$

Lemma 3 *The matrix P_t is symmetric and positive definite provided P_0 is positive definite.*

Proof. This is surely well known, but we provide a proof here for completeness. The proof is by induction. Since $var(\eta_t)$ is symmetric and positive definite, equation (6) shows it suffices to prove

$$P_{t-1} - P_{t-1} x_t D^{-1} x_t' P_{t-1}$$

is symmetric and positive definite, where $D = \sigma^2 + x_t' P_{t-1} x_t > 0$. Symmetry is trivial. By induction, the matrix inversion lemma may be applied to show

$$[P_{t-1} - K_t x_t x_t' P_{t-1}]^{-1} = P_{t-1}^{-1} + \frac{x_t x_t'}{\sigma^2}$$

thus showing $[P_{t-1} - K_t x_t x_t' P_{t-1}]^{-1}$ is positive definite and the result follows. ∎

Lemma 4 $[P_t - var(\eta_t)]^{-1} = P_{t-1}^{-1} + \frac{1}{\sigma^2} x_t x_t'$.

Proof. It is not obvious that $P_t - var(\eta_t)$ is invertible. However, since P_{t-1} is invertible, Lemma 2 applies to the expression $P_{t-1} - P_{t-1} x_t D x_t' P_{t-1}$ and the result follows by induction. ∎

We now transform the Kalman filter recursions so that the proof of Bray and Savin may be applied directly. To this end, set

$$R_t = \frac{\sigma^2}{t}\left[P_{t-1}^{-1} + \frac{1}{\sigma^2} x_t x_t'\right]$$

Note that by the above lemmas, R_t is symmetric and positive definite[6].

Lemma 5 $\frac{1}{t}R_t^{-1}x_t = K_t x_t$.

Proof. This is simply algebra. Notice, by Lemma 4

$$\frac{1}{t}R_t^{-1} = \frac{1}{\sigma^2}(P_t - var(\eta_t)) \tag{7}$$

which, by recursion (6), shows

$$\begin{aligned} \frac{1}{t}R_t^{-1}x_t &= \frac{1}{\sigma^2}\left[P_{t-1} - P_{t-1}x_t D^{-1}x_t' P_{t-1}\right]x_t \\ &= \frac{1}{\sigma^2}D^{-1}\left[P_{t-1}x_t\left[\sigma^2 + x_t'P_{t-1}x_t\right] - P_{t-1}x_t x_t' P_{t-1}x_t\right] \\ &= D^{-1}P_{t-1}x_t = K_t x_t. \end{aligned}$$

■

This lemma allows us to write equation (6) as

$$b_t = b_{t-1} + \frac{1}{t}R_t^{-1}x_t\left[y_t - b_{t-1}'x_t\right]. \tag{8}$$

Notice that if $R_t = \frac{1}{t}\sum x_i x_i'$ then equation (8) coincides with the RLS estimator of the linear model $y_t = \beta' x_t + \epsilon_t$. Furthermore, using the substitution $V_t = R_t^{-1}$, and plugging in the true data generating process (3), we see that recursion (8) is identical to recursion (2), up to the process V_t. As we mentioned previously, the proof of Bray and Savin's main result depended not on the specific process V_t, but only its almost sure convergence to $(Ex_t x_t')^{-1}$. Thus, to show almost sure convergence of the process (8) to $\frac{m}{1-a}$, it suffices to show R_t converges to $Ex_t x_t'$ almost surely.

Analysis of R_t is simplified using the following result.

Lemma 6 *The recursion for R_t may be written*

$$\rho_t(R_{t-1}, var(\eta_{t-1})) = -\frac{t(t-1)^2}{\sigma^2}R_{t-1}\left[\frac{t-1}{\sigma^2}R_{t-1} + var(\eta_{t-1})^{-1}\right]^{-1}R_{t-1} \tag{9}$$

$$R_t = R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) + \frac{1}{t^2}\rho_t(R_{t-1}, var(\eta_{t-1})). \tag{10}$$

Proof. Recall

$$R_t = \frac{\sigma^2}{t}\left[P_{t-1}^{-1} + \frac{1}{\sigma^2}x_t x_t'\right].$$

[6] This variable transform was used by Margaritis, [8]

Also, by (7),

$$P_{t-1}^{-1} = \left[\frac{\sigma^2}{t-1} R_{t-1}^{-1} + var(\eta_{t-1})\right]^{-1}.$$

Now simply apply Lemma 2 to complete the proof. ■

2.2.1 Back To Bray and Savin

Bray and Savin considered the case in which agents believed the parameters of the model to be constant. This is equivalent to the agents in our model believing that $var(\eta_t) = 0$. However, while Bray and Savin's agents used OLS to form their estimates, our agents use the Kalman filter. In this subsection, we show these estimators are equivalent.[7] This is not difficult. Indeed,

$$\lim_{\|var(\eta_{t-1})\|\to 0} \left(\rho_t(R_{t-1}, var(\eta_{t-1}))\right) =$$

$$\lim_{\|var(\eta_{t-1})\|\to 0} \left(\tfrac{t(t-1)^2}{\sigma^2} R_{t-1} var(\eta_{t-1}) \left[\tfrac{t-1}{\sigma^2} R_{t-1} var(\eta_{t-1}) + I\right]^{-1} R_{t-1}\right) = 0$$

The recursions defining the Kalman filter estimator, then, reduce to

$$\begin{aligned} y_t &= (m + \alpha b_{t-1})' x_t + u_t \\ b_t &= b_{t-1} + \frac{1}{t} R_t^{-1} x_t \left[y_t - b_{t-1}' x_t\right] \\ R_t &= R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}). \end{aligned}$$

These recursions are the same as those obtained by Bray and Savin and show that their model is a special case of the model we consider here.

2.2.2 The Bullard Result

In [2], Bullard showed, for a class of models which includes ours, that if agents believe the parameters of the perceived model to follow a random walk and if the conditional variance of the random walk is constant, then convergence to REE can not obtain. The idea behind this result is quite simple. For convergence to obtain (and to apply the main results of the theory of stochastic approximation) it must be the case that the gain of the algorithm goes to zero. In the Kalman filter recursions, this gain is represented by

$$K_t = P_{t-1}[\sigma^2 + x_t' P_{t-1} x_t]^{-1}.$$

[7] This type of equivalence was noted in a different model by Bullard, [2].

For this term to go to zero (almost surely) it must happen that P_t, as given by

$$P_t = P_{t-1} - K_t x_t x_t' P_{t-1} + var(\eta_t),$$

goes to zero (almost surely). But if $var(\eta_t) = Q > 0$, this can not happen. Note that P_t represents the agents' perceived mean square error at time t. If agents believe the parameters of the model will always have some non-zero constant conditional variance, they will always believe the MSE is non-zero, and in fact, bounded away from zero by the conditional variance of the parameters. Further, if their perceived MSE is strictly positive and bounded away from zero, agents will always be willing to adjust their estimates in the presence of forecast error; forecast error which will occur because of the stochastic nature of the model. Thus the agents estimators can not possibly converge to a constant value.

Also, we note here that if $var(\eta_t)$ does not converge to zero, then Bullard's result still holds. To show this, it suffices to show that, in this case, P_t does not converge to zero. Since $var(\eta_t)$ does not converge to zero, there is a subsequence, indexed, say, by $t(k)$, which is bounded away from zero. Since $P_t \geq var(\eta_t)$, it follows that $P_{t(k)}$ is bounded away from zero, and thus P_t can not converge to zero.[8]

2.2.3 Vanishing Variance

The results of the previous section indicate that a necessary condition for convergence to the REE is that the conditional variance of the random walk be decreasing to zero. And, as mentioned in the introduction, we believe this is a reasonable assumption to make, for if agents initially use OLS to estimate their parameters then Bray and Savin's result implies that the conditional variance of the parameters does decrease (eventually) since convergence of the agents' estimators to a constant value does obtain. In this section, we take as given that agents believe the conditional variance of the random walk is decreasing to zero and consider what rate is sufficient to guarantee convergence to the REE.

Recall the recursion describing R_t is given by

$$R_t = R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) - \frac{1}{t^2}\rho_t(R_{t-1}, var(\eta_{t-1})) \tag{11}$$

Further, we have seen that the form of the Kalman filter recursions, properly transformed, together with the proof of Bray and Savin's main result, shows that if $a < 1$ then convergence of b_t to $\frac{m}{1-a}$

[8] The notion of "greater than" for matrices is reviewed in Appendix A.

occurs with probability one provided R_t converges to Ex_tx_t' almost surely. Analysis of stochastic processes of the form (11) may be done using the theory of stochastic approximation. Under certain restrictions on the functions ρ_t, a differential equation may be analyzed to determine possible points of convergence. Unfortunately, due to the form of ρ_t, only local convergence results may be applied and the restrictions on the conditional variance are strong. This is discussed in detail in Section 3 of this paper. Fortunately, it is possible to prove global convergence (in the sense described below) with weaker restrictions, using a less direct approach. Specifically, we obtain the following result.

Lemma 7 *If R_1 is positive definite and symmetric, and*

$$limsup_t t^2 \|var(\eta_t)\| = 0 \tag{12}$$

then R_t converges to Ex_tx_t' almost surely.

The proof of this lemma is in Appendix A. The implication of restriction (12) is that the norm of the conditional variance must die just a little faster than $\frac{1}{t^2}$. Also, the convergence is global with respect to the initial condition subject to the restriction that the initial condition is symmetric, positive definite. Note that R_1 is symmetric, positive definite provided P_0 is, and that P_0 represents the agents' perceived mean square error of their initial belief, b_0. We conclude that this is not a significant restriction.

Lemma 7 together with the observations above concerning the application of Bray and Savin's proof to our recursions yields the following theorem which is the main result of this paper.

Theorem 8 *If P_0 is positive definite and*

$$limsup_t t^2 \|var(\eta_t)\| = 0$$

then $b_t \to \frac{m}{1-a}$ almost surely, provided $a < 1$.

Observe that the restriction on the parameter a, called the stability parameter, is the same as obtained by Bray and Savin. This is not surprising. The recursions describing the time path of b_t is the same for both models, except for the value of the positive definite matrix modifying the forecast error. Specifically, both recursions may be written as

$$b_t = b_{t-1} + \frac{1}{t} V_t x_t((1-a)b_{t-1}'x_t + m'x_t + u_t),$$

and the only difference will be the values of the positive definite matrix V_t. The restriction on a guarantees b_t moves toward its REE value. Because V_t is positive definite in both models, its specific value does not affect this direction.

The theorem predicts convergence obtains provided agents believe the conditional variance of the random walk eventually decreases a little faster than t^{-2}. Also, as Bullard showed, convergence does not obtain if the conditional variance is constant. What happens when the conditional variance decreases to zero at a rate less than or equal to t^{-2} has not been determined analytically. In a companion paper, [9], we report the results of simulations indicating convergence can obtain for rates of decrease slightly less than t^{-2} and also that for slow rates of decrease, say $t^{-.5}$, convergence does not appear to obtain.

3 A More General Cobweb Model

The analysis of stochastic recursive algorithms is usually done using the theory of stochastic approximation. For economic models, the standard is to use the results of Ljung, [6], Marcet and Sargent, [7], and Evans and Honkapohja, [4], which tell us to associate with the given stochastic process a differential equation. It can then be shown that, under certain conditions, possible convergence points of the process correspond to stable fixed points of the differential equation.[9] In this section, we apply the theory of stochastic approximation to the sequence of estimators obtained from a more general cobweb model. We obtain local convergence results provided stronger restrictions are placed on the rate of decrease of the conditional variance.

3.1 The Model

We generalize the model used by Bray and Savin to include serially correlated observable shocks. Specifically, we consider a reduced form Muth-model as given by

$$\begin{aligned} y_t &= aE^*_{t-1}y_t + \lambda' x_t && (13) \\ x_t &= Bx_{t-1} + \nu_t, \end{aligned}$$

[9]Note the power of this type of result. It tells us not only that convergence occurs, but also yields the possible limit points. And, being fixed points of differential equations, these limit points are often not difficult to compute.

where $x_t \in \mathrm{R}^n$ is an asymptotically stationary process with first component equal to one and $y_t \in \mathrm{R}$. For technical reasons, we require the *i.i.d.* process ν_t to be almost surely uniformly bounded. Deviating slightly from the assumptions of the previous model, we assume the expected time t value of the endogenous variable is formed with respect to information available at time $t-1$. This information includes x_{t-1}. This form of the model has been studied by, for example, Evans and Honkapohja, [4].

The unique rational expectations equilibrium for this model is given by

$$y_t = \frac{\lambda' B}{1-a} x_{t-1} + \lambda' \nu_t.$$

To incorparate learning into the model, we assume agents believe the final form of the model to be

$$\begin{aligned} y_t &= \beta_t' x_{t-1} + \varepsilon_t, \qquad (14) \\ \beta_{t+1} &= \beta_t + \eta_t. \end{aligned}$$

Notice this is the same form of beliefs as in the previous model, except for the timing of the exogenous variable.

Given the beliefs of the agents, the natural estimator is again given by the Kalman filter. Denote by b_t the Kalman filter estimate of β_{t+1} using information available at time t. Then, at time $t-1$, agents believe y_t will be determined by the following equation called the perceived law of motion, or PLM:

$$y_t = b_{t-1}' x_{t-1} + \varepsilon_t. \qquad (15)$$

Using the PLM, agents determine $E_{t-1}^*(y_t) = b_{t-1}' x_{t-1}$. This may be inserted into equation (14) to obtain the actual data generating process. Define the map $T : \mathrm{R}^n \to \mathrm{R}^n$ by

$$T(b) = ab + \lambda' B. \qquad (16)$$

Then the data is generated by the following equation called the actual law of motion, or ALM:

$$y_t = T(b_{t-1})' x_{t-1} + \lambda' \nu_t. \qquad (17)$$

The map T takes the perceived parameters to the actual parameters. When the perceived paramters equal the actual parameters, the model in is a rational expectations equilibrium; and thus a fixed point of the $T-map$ determines an REE.

The recursions describing the agents' estimators may now be reported. Since the state space model (15) describing agents beliefs is identical, up to timing, to the state space model considered earlier, the recursions describing the Kalman filter estimator are identical up to timing as well. We obtain

$$K_t = P_{t-1}[\sigma^2 + x'_{t-1}P_{t-1}x_{t-1}]^{-1} \tag{18}$$

$$b_t = b_{t-1} + K_t x_{t-1}[(T(b_{t-1}) - b_{t-1})' x_{t-1} + \lambda' \nu_t] \tag{19}$$

$$P_t = P_{t-1} - K_t x_{t-1} x'_{t-1} P_{t-1} + var(\eta_t) \tag{20}$$

where the actual law of motion has been inserted for y_t. As before, our goal is to analyze the asymptotic behavior of b_t.

3.2 E-stability

The notation developed above allows us to make an important link to the learning literature. To determine the stability under learning of a given REE, the industry standard is to use the theory of E-stability.[10] This theory tells us to consider the differential equation[11]

$$\frac{db}{d\tau} = T(b) - b. \tag{21}$$

Notice that an REE, b^*, is a stationary solution to this differential equation. The REE is said to be *E-stable* provided it is locally asymptotically stable. The *E-stability Principle* says that an E-stable REE is locally stable under learning provided a reasonable learning algorthim is employed. This intuition behind this principle is not difficult. Suppose that by reasonable learning algorthym it is meant that the new parameter estimates are obtained by moving in the direction of the forecast error redirected appropriately by the value of the regressor[12]. Given the actual law of motion, this product may be written

$$x_{t-1}x'_{t-1}(T(b_{t-1}) - b_{t-1}) + x_{t-1}\nu'_t\lambda.$$

Because $x_{t-1}x'_{t-1}$ is positive definite, the components of the vector determined by above expression will, on average, have the same signs as the components of $T(b_{t-1}) - b_{t-1}$. Thus, if the differential equation above is locally asymptotically stable at b^*, then moving according to the learning algorithm should, on average, result in convergence to b^*. It should be noted that the E-stability Principle is not a general result, and so, whenever possible, convergence should be proven using other techniques.

[10]Originally coined by George Evans.

[11]The $T - map$, and hence E-stability, may be defined for many different types of models. Also, the $T - map$ depends only on the reduced form equation and the agents' perceived beliefs and thus is independent of parameter estimation proceedure.

[12]Both the OLS estimator and the Kalman filter estimator behave in this manner.

Recall that a differential equation of the form (21) is locally asymptotically stable at b^* provided the eigenvalues of the derivative have real part less than zero. In our case, the derivative is $DT - I$ so local asymptotic stability requires that the eigenvalues of DT must have real part less than one. Since $DT = aI$ it follows that the REE is E-stable provided $a < 1$[13]. It is well known that $a < 1$ implies convergence to REE of the least squares learning algorithm: see, for example [4]. Thus, E-stability determines convergence when least squares learning is used. We will show that E-stability determines convergence when Kalman filter learning is used, provided the conditional variance of the random walk decreases rapidly.

3.3 Convergence

To analyze the asymptotic behavior of the agents' estimators we use the theory developed in Evans and Honkapohja [4]; for a summary, see [3]. Consider a recursive stochastic algorithm of the following form:

$$\begin{aligned} \theta_t &= \theta_{t-1} + \frac{1}{t} H(\theta_{t-1}, w_t) + \frac{1}{t^2} \rho_t(\theta_{t-1}, w_t) \qquad (22) \\ w_t &= A(\theta_{t-1}) w_{t-1} + B(\theta_{t-1}) \mu_t \end{aligned}$$

where μ_t is white noise. To this process is associated a differential equation as follows. Set

$$h(\theta) = lim_{t\to\infty} E\left(H(\theta, x_t)\right). \qquad (23)$$

The differential equation is

$$\frac{d\theta}{dt} = h(\theta). \qquad (24)$$

The main result of the theory says that if θ^* is a locally asymptotically stable fixed point of this differential equation, and H and ρ_t satisfy some nice properties in some neighborhood of this fixed point, then the process (22) converges to θ^* with probability one, provided a projection facility is used; see Appendix B for details.[14]

To employ the theory described above, we must put our algorithm into the form (22). This is done using the same variable substitution that was used to analyze the previous model. Specifically,

[13] Of course this is the same restriction on the stability parameter as obtained by Bray and Savin for their slightly different model

[14] Informally, a projection facility puts the process back near the fixed point if it wanders too far away. For details, see Appendix B. Projection facilities we introduced by Ljung, [6] and Marcet and Sargent, [7]. Evans and Honkapohja, [3][4], also show how weaker results can be obtained if the projection facility is dropped.

set

$$R_t = \frac{\sigma^2}{t}\left[P_{t-1}^{-1} + \frac{1}{\sigma^2}x_{t-1}x_{t-1}'\right].$$

Then the Kalman filter recursions may be rewritten as

$$\hat{\rho}_t(b_{t-1}, R_{t-1}) = -\frac{t(t-1)^2}{\sigma^2}R_{t-1}Q_{t-1}\left[\frac{t-1}{\sigma^2}R_{t-1}Q_{t-1} + I\right]^{-1}R_{t-1} \tag{25}$$

$$R_t = R_{t-1} + \frac{1}{t}(x_{t-1}x_{t-1}' - R_{t-1}) + \frac{1}{t^2}\hat{\rho}_t(b_{t-1}, R_{t-1}) \tag{26}$$

$$b_t = b_{t-1} + \frac{1}{t}R_t^{-1}x_{t-1}\left((T(b_{t-1}) - b_{t-1})'x_{t-1} + \lambda'\nu_t\right) \tag{27}$$

where $Q_t = var(\eta_t)$. The dependence of b_t on R_t forces us to make the standard variable change $S_{t-1} = R_t$. Letting $\theta_t = [b_t', S_t']'$ then allows us to write the Kalman filter recursions in the form (22)[15] where $w_t = [x_t', x_{t-1}', \nu_t]'$,

$$H(\theta_{t-1}, w_t) = \begin{bmatrix} S_{t-1}^{-1}x_{t-1}\left((T(b_{t-1}) - b_{t-1})'x_{t-1} + \lambda'\nu_t\right) \\ (x_t x_t' - S_{t-1}) \end{bmatrix}, \tag{28}$$

$$\rho_t(\theta_{t-1}, w_t) = \begin{bmatrix} 0 \\ -\left(\frac{t}{t+1}\right)\left[\frac{t^3}{\sigma^2}S_{t-1}Q_t\left[\frac{t}{\sigma^2}S_{t-1}Q_t + I\right]^{-1}S_{t-1} + x_t x_t' - S_{t-1}\right] \end{bmatrix}. \tag{29}$$

Having placed our stochastic algorithm in the correct form, we now consider its associated differential equation. It is easily computed to be

$$\frac{db}{d\tau} = S^{-1}M\left(T(b) - b\right)$$

$$\frac{dR}{d\tau} = M - S$$

where $M = \lim_{t\to\infty} Ex_t x_t'$ exists and is positive definite because x_t is asymptotically stationary.

Just as in the previous model, the Kalman filter estimator reduces to the OLS estimator as $\|Q_t\| \to 0$. Since Q_t is only present in the ρ_t term, it does not affect the functional form of the associated differential equation. Thus we are not surprised to find that the differential equation above is identical to the one obtained when least squares learning is modeled. And, as has been shown in the literature on least squares learning, the unique fixed point of the above system is $b^* = \frac{\lambda' B}{1-a}$ and $S = M$, and this fixed point is locally asymptotically stable provided $a < 1$.

According to the previous paragraph, to show convergence to REE, it suffices to restrict the model so that the technical conditions on H and ρ_t are satisfied. As mentioned, the form of H seen

[15] We are being a little sloppy here. Technically the theorems apply to vector processes θ, but in our case R is a matrix. The complication is avoided via application of the column operator which takes a matrix to the associated column vector. We suppress this for notational simplicity as is standard in the literature.

here is not new, it is identical to the H obtained from least squares learning models. In particular, that it satisfies the appropriate conditions is well known: see, for example, [4]. On the other hand, the form of ρ_t is new and thus the restrictions must be considered.

The result of Evans and Honkapohja requires ρ_t to be bounded in t by a simple function of w for all S in compact sets surrounding the fixed point. Specifically, we must show there is a U with $M \in U$ so that given compact K in U there are constants C and q so that for all t and for all $S \in K$

$$|\rho_t(S, w)| \leq C(1 + |w|^q).$$

It is easier to work with matrix norms. The following lemma, which is surely well known, allows us to do that[16].

Lemma 9 *Let $A \in \mathrm{R}^{n \times n}$, $x_n \in \mathrm{R}^n$ and $\|A\| = sup_{|v| \leq 1} |Av|$ be the usual matrix norm. Then*

1. $n||A||^2 \geq (col(A))^2$
2. $\|xx'\| \leq n^2 |x|^2$

Proof. 1: It can be shown that $\|A\| = \max_i \sum_j |a_{ij}|$. Thus

$$\begin{aligned} n\|A\|^2 &= n \max_i \sum_j |a_{ij}|^2 + \sum_{j \neq k} |a_{ij} a_{ik}| \\ &\geq n \max_i \sum_j |a_{ij}|^2 \geq \sum_{ij} |a_{ij}|^2 = (col(A))^2. \end{aligned}$$

2:

$$\begin{aligned} \|xx'\| &= \max_i \sum_j |x_i x_j| \leq \sum_{ij} |x_i x_j| \\ &\leq n^2 \max_{ij} |x_i x_j| = n^2 \max_i x_i^2 \leq n^2 |x|^2. \end{aligned}$$

■

For ρ_t to be bounded in t it is clear we must have the conditional variance decreasing like $\frac{1}{t^3}$. To this end, we make the following assumption:

$$sup_{t>0} t^3 \|Q_t\| = \sigma^2 k < \infty. \tag{30}$$

[16]By "surely well known" I mean that it is easier to prove than try to look up, but I deserve no credit for its discovery.

Now pick an open set U about M which is contained in the set of invertible matrices and let $K \subset U$ be compact. Let

$$C_1 = \max_{S \in K} (k\|S^2\| + \|S\|).$$

and $C = \sqrt{n} \max\{C_1, n^2\}$. We may now show that the above restriction bounds ρ_t appropriately. As in the first model, R_t and hence S_t is symmetric, positive definite provided R_0 is. Using this fact, together with Lemma 13 in the Appendix, we have

$$\begin{aligned} |\rho_t(S,x)| &\leq \sqrt{n}\|\rho_t(S,x)\| \leq \sqrt{n}\left(k\|S^2\| + \|xx'\| + \|S\|\right) \\ &\leq \sqrt{n}\left(C_1 + n^2|x|^2\right) \leq C(1+|x|^2). \end{aligned}$$

We may summarize the implications of the above observations in the following theorem.

Theorem 10 *If the economy is given by the model (14) and if agents have beliefs given by (15) together with the restriction on the conditional variance, (30), and if agents use the Kalman filter to obtain their expectations, and if the value of the stability parameter, a is less than one, then the economy will converge to the REE with probability one provided the learning algorithm is augmented with a projection facility.*

4 Conclusion

Since its conception, the assumption of rational agents has been criticized as being too strong. The landmark paper of Bray and Savin, and the learning literature published since, has given credence to the rational expectations hypothesis because it has shown that, for many models, weakening the assumption to that of boundedly rational agents preserves rational expectations equilibria asymptotically. However, these least squares learning models have plagued by the same criticism originally borne by the simple adaptive models which predated and, in fact, led to the rational expectations hypothesis: why would reasonable agents make systematic errors?[17]

In this paper we have begun the process of addressing the issue of model misspecification. We have allowed our agents to increase their sophistication by postulating a time-varying process for the parameters of the model. We have shown that, for certain restrictions on the postulated process, convergence to REE still obtains. This further strengthens the learning literature's justification

[17]The error in the models of least squares learning, the misspecification of the model, is, admittedly, more subtle than before, but still sometimes detectable using standard econometric techniques.

for continued analysis of REE. And these results are in contrast to the non-convergence result obtained by Jim Bullard. He shows that if the conditional variance of the random walk is constant then convergence can not possibly obtain. We have argued that it is more natural to consider the specification that the conditional variance decreases to zero and subsequently over-turned his result.

The results of this paper are not complete. Several important questions remain. First: what is the asymptotic behavior of the economy if the conditional variance of the random walk decreases to zero more slowly that the restrictions required for our results? Simulations which are reported in a companion paper suggest that convergence does obtain for rates of decrease which are slower, but not all such rates. Second, what is effect of altering the assumption that the parameters follow a random walk? Bray and Savin suggest a return to normalcy process; such a process with constant conditional variance is analyzed by Bullard and he again obtains a non-convergence result. Analysis of such a model with decreasing variance has proven difficult because the form of the resulting recursions are not addressed by the stochastic approximation literature. Simulations reported in the companion paper suggest that convergence to REE obtains provided the conditional variance decreases rapidly enough and that the restrictions on this rate of decrease may be weaker than the analogous restrictions in the random walk model. Finally, our agents, just like Bray and Savin's, have a misspecified model. Can the agents use some natural econometric technique to detect this misspecification? We are working on this problem currently.

Appendix A: The Convergence of R_t

The argument that follows will repeatedly use the probabilistic notion of *event*. We employ the following notation: let x be a random variable and P be some property which realizations of x may or may not have. The $E = \{x \text{ has } P\}$ means E is the event that the realization of x has the property P. Further, if E and F are events then $E \subset F$ means the event F occurs whenever the event E occurs. Also, for a given random variable w, denote by $\tilde{w}$ a particular realization. Notice if $E = \{x \text{ has } P\}$ and $\tilde{x}$ has P then E has occured.

The following results concern almost sure convergence of stochastic processes. Recall that the process x_n converges to x almost surely (that is, with probability one) provided the event that the sequence of realizations of x_n converges to the realization of x occurs with probability one. In the sequel we make repeated use of the following lemma which is surely well known.

Lemma 11 *The sequence of random variables x_n converges almost surely to the random variable x if and only if for any $\epsilon > 0$ and any $p \in (0,1)$ there is an N so that*

$$Pr\{\|x_n - x\| \leq \epsilon \;\forall n \geq N\} > p.$$

Proof. (Necessity) Since $x_n \to x$ almost surely, it follows that $\|x_n - x\| \to 0$ almost surely. let $\epsilon > 0$ and $p \in (0,1)$ and for any N define the event $A(N)$ as follows:

$$A(N) = \{\|x_n - x\| < \epsilon \;\forall n \geq N\}.$$

Set $A = \cup_N A(N)$. Let E be the event that the realizations of x_n converge to the realization of x. Suppose E occurs. Then there is some M so that $n > M \Rightarrow \|\tilde{x}_n - \tilde{x}\| < \epsilon$ which implies the event $A(M)$ occurs. Thus $E \subset A$ and so $Pr(A) = Pr\left(\cup_N A(N)\right) = 1$. A well known result from measure theory tells us that if $M > N$ implies $A(N) \subset A(M)$ then

$$Pr\left(\cup_N A(N)\right) = \lim_N Pr(A(N)).$$

Since the inclusion requirement holds in our case, we see $\lim_N Pr(A(N)) = 1$ which proves necessity. (Sufficiency) Choose an increasing sequence $p_j \in (0,1)$ so that $p_j \to 1$ and a decreasing sequence $\epsilon_k > 0$ so that $\epsilon_k \to 0$. Let $N(p_j, \epsilon_k)$ be so that $Pr(A_{N(p_j,\epsilon_k)}) > p_j$ where

$$A_{N(p_j,\epsilon_k)} = \{\|x_n - x\| \leq \epsilon_k \;\forall n \geq N(p_j, \epsilon_k)\}.$$

Set

$$A = \bigcap_k \bigcup_j A_{N(p_j,\epsilon_k)}.$$

Let $\tilde{x}_n$ and $\tilde{x}$ be realizations and suppose A occurs. I claim E occurs, that is, $\tilde{x}_n \to \tilde{x}$. Let $\epsilon > 0$ and pick k so that $\epsilon_k < \epsilon$. Since A occured, so too did $\cup_j A_{N(p_j,\epsilon_k)}$ and so there is some j so that $A_{N(p_j,\epsilon_k)}$ occured. Then $n > N(p_j, \epsilon_k) \Rightarrow \|\tilde{x}_n - \tilde{x}\| < \epsilon$ which shows E occured. To complete the proof, we must show that the measure of A is one. Fix k. Then it suffices to show the probability of $\cup_j A_{N(p_j,\epsilon_k)}$ is one. Let $p < 1$. Choose j so that $p_j > p$. Then $Pr(\cup_j A_{N(p_j,\epsilon_k)}) > Pr(A_{N(p_j,\epsilon_k)}) > p$.
■

We say a matrix is positive if it is symmetric and positive definite, and non-negative if it is symmetric and positive semi-definite. If A and B are symmetric, we say $A < B$ if $B - A$ is positive (or written $B - A > 0$) and we say $A \leq B$ if $B - A$ is non-negative (or written $B - A \geq 0$.) These relations induce partial orderings on the set of symmetric matrices. It is well known that a matrix is positive (non-negative) provided all eigenvalues are positive (non-negative and real).

The recursion R_t is defined as follows. Let $x_t \in \mathrm{R}^n$ be *i.i.d.*. Then

$$\begin{aligned}
\rho_t(R_{t-1}, Q_{t-1}) &= \frac{(t-1)^2}{\sigma^2} R_{t-1} Q_{t-1} \left[\frac{t-1}{\sigma^2} R_{t-1} Q_{t-1} + I \right]^{-1} R_{t-1} \\
R_t &= R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) - \frac{1}{t} \rho_t(R_{t-1}, Q_{t-1}).
\end{aligned}$$

We assume that Q_t is positive for all t. We take the initial condition of the recursion to be a positive matrix and recall the previously mentioned implication that with probability one, all elements of the sequence R_t will be positive. Assume $\limsup_t t^2 \|Q_t\| = 0$ and set $E(x_t x_t') = \Omega$. Our goal is to prove the following result.

Theorem 12 *If R_0 is positive then the process R_t converges to Ω almost surely.*

This will be facilitated by the following lemma on positive matrices and the induced partial ordering.

Lemma 13 *Let A, B, and C be positive conformable matrices.*

1. *If $A \leq B$ then $AC \leq BC$.*
2. *If $A \leq B$ then $B^{-1} \leq A^{-1}$.*
3. *$A \leq \|A\| I$.*
4. *If $A \leq B$ then $\|A\| \leq \|B\|$.*

These results are standard. The first three may be found in [5]. To prove the fourth, proceed as follows: Statement 3 implies $A \leq \|B\| I$. Note that λ is an eigenvalue of A if and only if $\|B\| - \lambda$

is an eigenvalue of $\|B\|I - A$. Also, since A is symmetric, $\|A\| = \max_i \|\lambda_i\|$ where λ_i varies over the eigenvalues of A. Let λ_m be the eigenvalue of maximum modulus. Note that it is also real and positive. Since $\|B\|I - A \geq 0$, its eigenvalues must be non-negative. Thus $\|B\| - \|A\| = \|B\| - \lambda_m \geq 0$.

The following result relating the partial ordering on symmetric matrices to convergence is needed.

Lemma 14 *Suppose x_n is a sequence of stochastic matrices which are almost everywhere positive and which converge almost surely to the positive matrix x. Fix some $\delta > 0$ and for given N define the event E as*

$$E(N) = \{x_n \leq (\|x\| + \delta)I \ \forall n \geq N\}.$$

Then for any $p \in (0,1)$, there exists N so that $Pr\{E(N)\} > p$.

Proof. Define the event F as follows:

$$F(N) = \{\|x_n\| \leq \|x\| + \delta \ \forall n \geq N\}.$$

Because $x_n \to x$ a.s. we make choose N so that $Pr\{F(N)\} > p$. Now notice that whenever $F(N)$ occurs,

$$x_n \leq \|x_n\| I \leq (\|x\| + \delta)I, \ \forall n \geq N$$

which implies $F(N) \subset E(N)$. ■

The next lemma will be used to provide an upper bound for the sequence R_t. Define a new sequence as follows:

$$\hat{R}_t = \hat{R}_{t-1} - \frac{1}{t}(x_t x_t' - \hat{R}_{t-1}).$$

Note that $\hat{R}_t = \frac{1}{t}\sum_{i=1}^{t} x_i x_i'$ so that, by the law of large numbers, $\hat{R}_t$ converges to Ω almost surely.

Lemma 15 *$R_t \leq \hat{R}_t$.*

Proof. The proof is by induction. First notice $\rho_t(R_{t-1}, Q_{t-1})$ is positive. Thus

$$R_1 = x_1 x_1' - \rho_1 \leq x_1 x_1' = \hat{R}_1.$$

Assume $R_{t-1} \leq \hat{R}_{t-1}$. Then

$$\begin{aligned} R_t &= R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) - \frac{1}{t}\rho_t(R_{t-1}, Q_{t-1}) \\ &= \frac{t-1}{t}R_{t-1} + \frac{1}{t}x_t x_t' - \frac{1}{t}\rho_t(R_{t-1}, Q_{t-1}) \\ &\leq \frac{t-1}{t}\hat{R}_{t-1} + \frac{1}{t}x_t x_t' = \hat{R}_t. \end{aligned}$$

■

Now that we have established an upper bound for R_t we begin work on the lower bound. Fix $\delta > 0$ and let $\xi = \|\Omega\| + \delta$, and $M(T) = \sup_{t>T} \frac{t^2}{\sigma^2}\|Q_t\|$. Notice that, by assumption, $M(T) \to 0$ and is decreasing. For each T we define a new sequence as follows:

$$S_t(T) = \begin{cases} R_t & \text{if } t \leq T \\ S_{t-1}(T) + \frac{1}{t}(x_t x_t' - S_{t-1}(T)) - \frac{1}{t}M(T)\xi^2 I & \text{else} \end{cases}$$

It is simple to show that $S_t(T)$ converges to $\Omega - M(T)\xi^2 I$ almost surely. Indeed, a straightforward induction argument shows $S_t(T) = \frac{1}{t}\sum_{i=1}^t x_i x_i' - M(T)\xi^2 I$, and the law of large numbers completes the result. The following lemma relates these sequences to R_t.

Lemma 16 *Let $p \in (0,1)$ and*

$$E(T) = \{S_t(T) \leq R_t \ \forall t\}.$$

Then there exists a $\hat{T}$ so that $T \geq \hat{T}$ implies $Pr(E(T)) > p$.

Proof. Set

$$F(T) = \{\hat{R}_t \leq \xi I \ \forall t \geq T\}.$$

By Lemma 14 there is a $\hat{T}$ so that $Pr(F(\hat{T})) > p$. To complete the proof then, it suffices to show that $T_2 \geq T_1 \Rightarrow F(T_1) \subset E(T_2)$. The proof is by induction. Assume $F(T_1)$ occurs. By construction $S_t(T_2) \leq R_t$ for $t \leq T_2$. So let $t > T_2$ and assume $S_{t-1}(T_2) \leq R_{t-1}$. Since $F(T_1)$ occured and $t > T_2 > T_1$ it follows, using Lemma 13, that

$$\begin{aligned} \rho_t(R_{t-1}, Q_{t-1}) &= \frac{(t-1)^2}{\sigma^2} R_{t-1} Q_{t-1} \left[\frac{t-1}{\sigma^2} R_{t-1} Q_{t-1} + I\right]^{-1} R_{t-1} \\ &\leq M(T_2) R_{t-1} \left[\frac{t-1}{\sigma^2} R_{t-1} Q_{t-1} + I\right]^{-1} R_{t-1} \\ &\leq M(T_2) R_{t-1}^2 \leq M(T_2)\xi^2 I. \end{aligned}$$

It follows that

$$\begin{aligned} S_t(T_2) &= S_{t-1}(T_2) + \frac{1}{t}(x_t x_t' - S_{t-1}(T_2)) - \frac{1}{t}M(T)\xi^2 I \\ &= \frac{t-1}{t} S_{t-1}(T_2) + \frac{1}{t}x_t x_t' - \frac{1}{t}M(T)\xi^2 I \\ &\leq \frac{t-1}{t} R_{t-1} + \frac{1}{t}x_t x_t' - \frac{1}{t}M(T)\xi^2 I \\ &\leq \frac{t-1}{t} R_{t-1} + \frac{1}{t}x_t x_t' - \frac{1}{t}\rho_t(R_{t-1}, Q_{t-1}) = R_t. \end{aligned}$$

■

We are now ready to prove the main result.

Proof. Let $p \in (0,1)$ and $\epsilon > 0$. As usual, we have

$$\|R_t - \Omega\| \leq \|R_t - S_t(T)\| + \|S_t(T) - \hat{R}_t\| + \|\hat{R}_t - \Omega\|.$$

Set

$$D(K) = \{\|\hat{R}_t - \Omega\| < \frac{\epsilon}{3} \ \forall t \geq K\}$$

and choose $\tilde{K}$ so that $Pr(D(\tilde{K})) > \frac{2+p}{3}$. Choose T_1 so that $M(T_1)\xi^2 < \frac{\epsilon}{6}$ and choose $T_2 > T_1$ so that $Pr(E(T_2)) > \frac{2+p}{3}$ where $E(T)$ is the event as defined in Lemma 16. This fixes the sequence $S_t(T_2)$. Define the event F by

$$F = \{0 \leq \|R_t - S_t(T_2)\| \leq \|\hat{R}_t - S_t(T_2)\| \ \forall t\}.$$

Notice that, by statement four of Lemma 13, $E(T_2) \subset F$.

Since $S_t(T_2) \to \Omega - M(T_2)\xi^2 I$ and $\hat{R}_t \to \Omega$ almost surely, it follows that $\|\hat{R}_t - S_t(T_2)\| \to M(T_2)\xi^2$ almost surely. Let

$$G(K) = \{\|\hat{R}_t - S_t(T_2)\| < M(T_2)\xi^2 + \frac{\epsilon}{6} \ \forall t \geq K\}.$$

Then we may choose $\hat{K} > \tilde{K}$ so that $Pr(G(\hat{K})) > \frac{2+p}{3}$.

Now let $H(K)$ be the following event:

$$H(K) = \{\|R_t - \Omega\| < \epsilon \ \forall t \geq K\}.$$

The main result is proved by showing we may choose K so that $Pr(H(K)) > p$. We claim $\hat{K}$ suffices. To see this, first notice $G(\hat{K}) \cap F \cap D(\tilde{K}) \subset H(\hat{K})$. Indeed, suppose $G(\hat{K}) \cap F \cap D(\tilde{K})$ occurs. Then, for all $t > \hat{K}$

$$\begin{aligned} \|R_t - \Omega\| &\leq \|R_t - S_t(T_2)\| + \|S_t(T_2) - \hat{R}_t\| + \|\hat{R}_t - \Omega\| \\ &\leq M(T_2)\xi^2 + \frac{\epsilon}{6} + M(T_2)\xi^2 + \frac{\epsilon}{6} + \frac{\epsilon}{3} < \epsilon. \end{aligned}$$

Finally, note

$$\begin{aligned}
Pr(G(\hat{K}) \cap F \cap D(\tilde{K})) &= 1 - Pr\left(\left(G(\hat{K}) \cap F \cap D(\tilde{K})\right)^c\right) \\
&= 1 - Pr\left(G(\hat{K})^c \cup F^c \cup D(\tilde{K})^c\right) \\
&\geq 1 - \left[Pr(G(\hat{K})^c) + Pr(F^c) + Pr(D(\tilde{K})^c)\right] \\
&\geq 1 - 3\left[1 - \frac{2+p}{3}\right] = p
\end{aligned}$$

■

Appendix B: Results on Stochastic Approximation

The main result of Section 3 was proved using a theorem by Evans and Honkapohja. We present the remaining details of that theorem's use here. We placed our algorithm in the required form, repeated here for convenience:

$$\theta_t = \theta_{t-1} + \frac{1}{t} H(\theta_{t-1}, w_t) + \frac{1}{t^2} \rho_t(\theta_{t-1}, w_t).$$

To apply the results of Evans and Honkapohja, it must be shown that H and ρ_t satisfy certain conditions. These conditions are carefully reported in [4]. The form of H in our model is not new; it is well known, and shown in [4], that H has the appropriate properties. The form of ρ_t was shown, in Section 3, to have the appropriate properties, provided the conditional variance decreased to zero rapidly enough. We may now proceed to state and apply the result of Evans and Honkapohja.

The differential equation, (21) is locally asymptotically stable at (b^*, M) provided $a < 1$. Let D be an open domain of attraction of (b^*, M). By the converse to Lyapunov's theorem (see [4], Proposition 5.9) there is a Lyapunov function $U : D \to \mathrm{R}$ with the following properties:

1. $U(b^*, M) = 0$ and $U(\theta) > 0$ for all $\theta \neq (b^*, M)$
2. $U(\theta) \to \infty$ as $\theta \to \partial D$.

For $c > 0$ set $K(c) = \{\theta : U(\theta) < c\}$. Note that $c_1 \leq c_2 \Rightarrow K(c_1) \subset K(c_2) \subset D$. For fixed c_2 and $c_1 < c_2$ define the projection facility as follows:

$$P(\theta) = \begin{cases} \theta & \text{if } \theta \in K(c_2) \\ \in K(c_1) & \text{if } \theta \notin K(c_2) \end{cases}$$

The theorem of Evans and Honkapohja allows us to conclude that if the algorithm is augmented with the above projection facility, then the process $\theta_t = \begin{bmatrix} b_t \\ S_t \end{bmatrix}$ converges almost surely to (b^*, M).

References

[1] M. Bray and N. Savin. Rational expectations, learning, and model specification. *Econometrica*, pages 1129 – 1160, 1986.

[2] J. Bullard. Time-varying parameters and non-convergence to rational expectations equilibrium under least squares learning. *Economics Letter*, pages 159 – 166, 1992.

[3] G. Evans and S. Honkapohja. Learning dynamics. In J. Taylor and M. Woodford, editors, *Handbook of Macroeconomics*. Elsevier, Amsterdam, 1998.

[4] G. Evans and S. Honkapohja. *Learning and Expectations in Macroeconomics*. Princeton University Press, 2000.

[5] R. Kadison and J. Ringrose. *Fundamentals of the Theory of Operator Algebras, Volume 1*. Academic Press, 1983.

[6] L. Ljung. Analysis of stochastic recursive algorythms. *IEEE Transactions on Automatic Control*, pages 551 – 575, 1977.

[7] A. Marcet and T. Sargent. Convergence of least squares mechanisms in self-referential linear stochastic models. *Journal of Economic Theory*, pages 337–368, 1989.

[8] D. Margaritis. A time=varying model of rational learning. *Economics Letter*, pages 309 – 314, 1990.

[9] B. McGough. A numerical study of learning models with time-varying parameters. *Working Paper*, Forthcomming.